ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

Automatic Pain Intensity Recognition: Training Set Selection based on Outliers and Centroids

Peter Bellmann[1], Patrick Thiam[2,1] and Friedhelm Schwenker[1] *

1- Ulm University - Institute of Neural Information Processing
James-Franck-Ring, 89081 Ulm - Germany

2- Ulm University - Institute of Medical Systems Biology
Albert-Einstein-Allee 11, 89081 Ulm - Germany

Abstract. In this study, we evaluate a person independent pain intensity recognition task, based on the BioVid Heat Pain Database. Previous works show that for such classification tasks, the overall performance can be increased by reducing the training data, based on certain criteria, such as different distance measures. This results in considering only a certain amount of participants from the training set, whose data distributions are defined to be the most similar to the data distribution of the participant from the test set. Counterintuitively, we propose to remove participants, which are identified as central points, from the training set, completely independent from the test set. Our evaluations show that this approach can lead to significant improvement of classification accuracy.

1 Introduction

Pain intensity recognition is still a challenging task in the field of e-health. Previous studies show that the overall performance of a classification model can be increased, if the classification system is trained on a specific subset of the available training data. The special situation in data sets such as the BioVid Heat Pain Database is that the data is organised in subject subsets. Different authors propose selecting specific training subsets, based on the *similarity* between the participants, which represent the whole training set and the person representing the test set (see Sec. 3). We propose to exclude data subsets specific to participants, solely based on the fully available training set, without any knowledge of the test data distribution. Our experiments show that the removal of data specific to one single participant is already sufficient for a significant improvement of the classification model's generalisation ability. In this study, we focus on distance based training set selection. Interestingly, our experimental outcomes support our idea that keeping participants, which are identified as *central points*, in the training set seems to harm the overall classification performance.
This study is organised as follows. In Sec. 2, we describe the BioVid Heat Pain Database. Section 3 provides a couple of related works, as well as the motivation for our approach. In Sec. 4, we define our validation protocol. The outcomes are presented and discussed in Sec. 5. Finally, in Sec. 6, we conclude this study.

*We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research. The work of Peter Bellmann and Friedhelm Schwenker is supported by the project *Multimodal recognition of affect over the course of a tutorial learning experiment* (SCHW623/7-1) funded by the German Research Foundation (DFG).

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

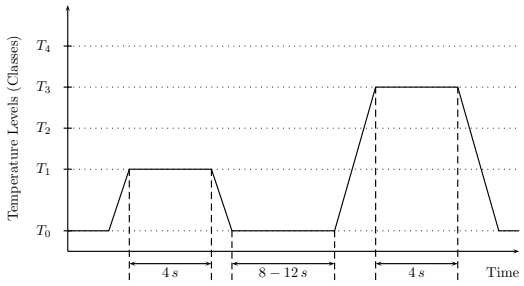

Fig. 1: An example sequence for a participant's stimuli and recovering phases.

2 BioVid Heat Pain DataBase

In this study, we focus on the publicly available BioVid Heat Pain Database (BVDB) [1], which was collected at Ulm University for pain intensity and emotion recognition research purposes. We use part A[1] of the BVDB, which comprises 87 participants, with focus on the recorded biopotentials for the pain intensity recognition task. Pain was elicited locally at the participant's forearm, by strictly controlled heat stimuli, using a Medoc thermode[2]. The *neutral state* temperature was defined as $32°C$ (T_0). The first part of the data acquisition experiments included an individual calibration phase, for each participant, which was undertaken to define four equidistant pain intensity levels (T_1, T_2, T_3, T_4).

Pain stimuli. The participants were stimulated twenty times with each of the pain related temperature levels, in randomised order for a duration of four seconds. Between two pain stimuli, the participants were stimulated with the neutral state temperature, with a stimulation length randomised from eight to twelve seconds. Figure 1 depicts an example of a temperature stimuli sequence.

Recorded biopotentials. Three different physiological channels have been recorded, including electrocardiogram (ECG) that measures heart activity, electrodermal activity (EDA) that measures skin conductance, and electromyogram (EMG) that measures muscle activity. The EDA signals were recorded at the participants' index and ring fingers. The EMG sensors recorded activity of the trapezius muscle, which is located in the upper back area of a human torso.

Feature extraction. For each sample, the physiological features were extracted from windows of length 5.5 sec. This was especially important for the samples specific to the pain related temperature levels. To reduce artefacts and noise in the physiological signals, different smoothing and signal detrending techniques were applied. Then, different statistical descriptors, such as *mean*, *standard deviation*, *minima* and *maxima*, etc., were extracted from the temporal domain. Moreover, additional features, including *bandwidth*, *central* and *mean frequency*, etc., were extracted from the frequency domain. Finally, the whole procedure led to 194 physiological features, in total. We refer the readers to [2] for a detailed description on the applied feature extraction and normalisation.

[1]More details on http://www.iikt.ovgu.de/BioVid.print
[2]For full information, see https://medoc-web.com/products/pathway/

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

3 Related Work & Motivation

In [2], Kächele et al. showed that the overall accuracy can be improved, based on the test subjects' data distribution, by finding *similar* participants, i.e. *nearest neighbours*, and training the classification model solely on those neighbours instead of on the fully available training data. Thiam et al. came to the same conclusion, in [3], based on a similar data set.
As we mentioned in Sec. 1, the data is organised in subject subsets (see also Sec. 2). Therefore, by the term *participant* we will denote the *data subset* specific to the participant, throughout the rest of this study. One basic feature of data preprocessing is the removal of outliers, which are defined as data points that are *far away* from all other data points. In this work, we additionally evaluate a counterintuitive approach, by analysing the effects of removing *centroids* from the data. Analogous to the definition of outliers, we define a centroid as the data point, which is *near* to *many* other data points, i.e. which has the lowest sum of distances to all other data samples (see Sec. 4). Therefore, both, outliers as well as centroids, define extreme cases, in each data set. However, it is common to remove only outliers from the (training) data.

4 Experimental Settings

Let $X \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be a d-dimensional (training) set. Furthermore, let $N \in \mathbb{N}$ be the number of participants whose samples constitute X. For each $x \in X$, we define $p(x)$ as the participant ID, i.e. $p : X \mapsto \{1, \dots, N\}$. Moreover, by $\bar{x}_i \in \mathbb{R}^d$, we denote the mean-based prototype of the data distribution of participant i, i.e. the d-dimensional vector including the mean values for each feature. Our proposed outlier and centroid detection is based on the sums of participant specific distances $d, \bar{d} \in \mathbb{R}^N$, which we define as follows,

$$d_i := \sum_{\substack{x \in X \\ p(x)=i}} \sum_{\substack{y \in X \\ p(y) \neq i}} \|x - y\|_2, \qquad \bar{d}_i := \sum_{j=1}^{N} \|\bar{x}_i - \bar{x}_j\|_2, \qquad \forall\, i = 1, \dots, N. \quad (1)$$

In each cross validation run of our initial experiments (see Sec. 5), we determine one single participant from the current set, which will be removed from the training set. Thereby, we compare four different evaluation approaches. In the first two approaches, we remove the participant, which corresponds to the minimum d_i and $\bar{d}_i$ values, denoted by MIN and $\overline{MIN}$ respectively. In the latter two approaches, we remove the participant, which corresponds to the maximum d_i and $\bar{d}_i$ values, denoted by MAX and $\overline{MAX}$ respectively. Therefore, participants detected by MIN and $\overline{MIN}$ are defined as centroids, whereas participants detected by MAX and $\overline{MAX}$ are defined as outliers. Table 1 summarises the experimental settings, which are applied in this study.

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

Table 1: Summary of experimental settings. LOPO CV: Leave-One-Participant-Out Cross Validation. RF: Random Forest (number of base classifiers) [4]. Y: Test Set.

Evaluation	Classification Model	Performance Measure
LOPO CV	RF (500 Trees)	$\lvert\{y \in Y : y \text{ classified correctly}\}\rvert / \lvert Y \rvert$

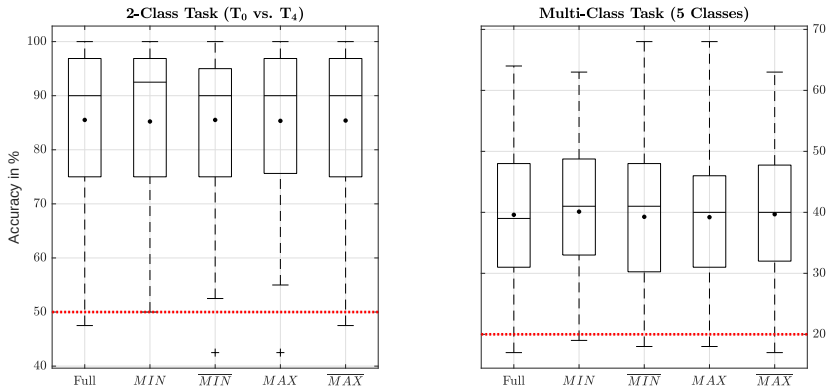

Fig. 2: Leave-One-Participant-Out cross validation performance values. The mean and the median values are represented by a dot and a horizontal line, respectively. The dotted (red) lines denote the chance level accuracies (20% in a 5-class classification task). Full: Training on the fully available training set. $MIN/\overline{MIN}/MAX/\overline{MAX}$: Removal of one participant corresponding to $\min d/\min \bar{d}/\max d/\max \bar{d}$ (d and $\bar{d}$ are defined in Eq. (1)) from the training set.

5 Experimental Validation

In the first part of this section, we provide the results based on the settings from Sec. 4, for the whole BVDB, as well as for the 2-class classification subset of the BVDB, which is defined by the T_0 vs. T_4 task. Subsequently, based on the outcomes of the initial experiments, we add further experimental evaluations for the multi-class task, in which all five available classes are considered.

Initial Results

Figure 2 depicts the results for the experimental settings, which are discussed in Sec. 4 (see also Table 1), for the removal of one single participant from the training set. The results based on the 2-class classification task (left part of Fig. 2) show that there is no significant difference between the proposed training set selection techniques, when only one participant is defined to be removed from the training set. However, the best median value is achieved by the method, which we denote by MIN (centroid detection based on all available data points).

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

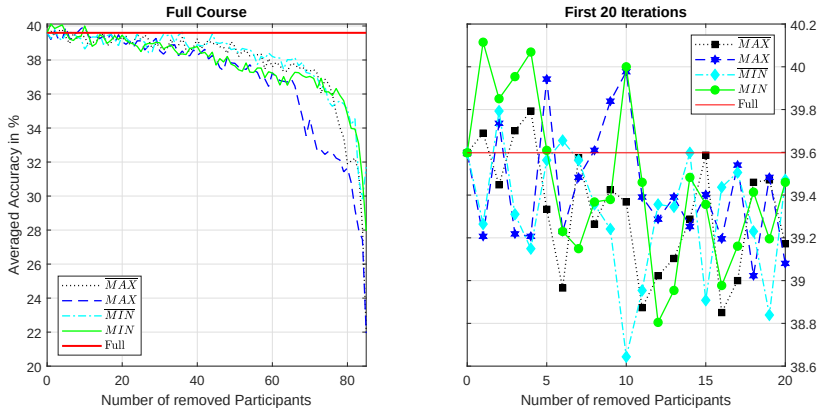

Fig. 3: Leave-One-Participant-Out cross validation performance values. Full: Training on the fully available training set. $\overline{MAX}/MAX/\overline{MIN}/MIN$: Removal of participants corresponding to $\max \bar{d}/\max d/\min \bar{d}/\min d$ (d and $\bar{d}$ are defined in Eq. (1)) from the training set.

In the multi-class task, the results vary more. From the right part of Fig. 2, we can make the following observations. Determining one participant that is removed from the training set, based on the minimum sum of distances works better when all available data points are considered. In contrast, determining one participant based on the maximum sum of distances leads to better performance values when only the prototypes of participants are considered. Moreover, applying the two-sided Wilcoxon signed-rank test [5], at a significance level of 5%, implies that the approach, which we denote by MIN is the only one leading to a significant improvement ($p = 0.0378$), in comparison to using the fully available training set. Note that the significant drop in accuracy, by changing the task from two classes to five, is due to the *complexity* of the BVDB. Similar results are reported, e.g. in [2] and [6], including a short discussion on the complexity of the BVDB. Moreover, based on the BVDB, automatic feature extraction based on deep physiological models is discussed in [7].

Follow-up Results

In this part of the experiments, instead of removing one single participant, we provide the results for removing $k = 1, \ldots, N-1$ participants for all four approaches, for the multi-class task. From the left part of Fig. 3, we can make the following observations. As expected, a *strong* reduction of the training set leads to a significant drop in accuracy. Removing a *huge* amount of participants (> 40) according to the method, which we denote by MAX (outliers based on all data points), leads to the worst results. However, training the classification model on one single participant still outperforms the chance level (classification performance: 20%) significantly for all proposed approaches.

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

From the right part of Fig. 3, we can make the following observations. Similar to the removal of centroids, the achieved mean accuracy values are better when the outliers are defined based on all data samples (MAX), instead of on prototypes ($\overline{MAX}$). The best overall result is achieved by the MIN approach when one single participant is removed from the training set. Moreover, the top three results were obtained by the MIN approach (for the removal of 1, 4 and 10 participants from the training set).
Note that the BVDB has the special property that each participant has exactly the same amount of data samples. For imbalanced data sets, where each participant is represented by a different number of data samples, one has to find an appropriate weighting factor for the sum of distances in the left part of Eq. (1), e.g. one divided by the number of distance values.

6 Conclusion

In this study, we introduced our idea of defining the participant with the lowest sum of distances to all other participants, i.e. the central point, as a candidate for training data clean up. For future work, we define the following research directions. First, one should test our approach, which we denoted by MIN, on other affect related data sets to confirm its effectiveness. Second, in this work we evaluated four different distance based approaches for participant based outlier and centroid detection. It could also be beneficial to combine (the best) two of the proposed methods. And third, instead of compressing the data specific to one participant to one single prototype ($\bar{d}$ from Eq. (1)), one could compute class-specific prototypes for each participant to preserve some information of each participant's data distribution. In general, the reported outcomes could motivate the implementation of novel approaches for data preprocessing/training set selection techniques, based on the removal of centroids.

References

[1] Steffen Walter, Sascha Gruss, Hagen Ehleiter, Jun-Wen Tan, Harald C. Traue, Stephen Clive Crawcour, Philipp Werner, Ayoub Al-Hamadi, and Adriano O. Andrade. The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system. In *CYBCONF*, pages 128–131. IEEE, 2013.

[2] Markus Kächele, Patrick Thiam, Mohammadreza Amirian, Friedhelm Schwenker, and Günther Palm. Methods for person-centered continuous pain intensity assessment from bio-physiological channels. *J. Sel. Topics Signal Processing*, 10(5):854–864, 2016.

[3] Patrick Thiam, Viktor Kessler, and Friedhelm Schwenker. Hierarchical combination of video features for personalised pain level recognition. In *ESANN*, pages 465–470, 2017.

[4] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[5] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[6] Peter Bellmann, Patrick Thiam, and Friedhelm Schwenker. *Multi-classifier-Systems: Architectures, Algorithms and Applications*, pages 83–113. Springer International Publishing, Cham, 2018.

[7] Patrick Thiam, Peter Bellmann, Hans A. Kestler, and Friedhelm Schwenker. Exploring deep physiological models for nociceptive pain recognition. *Sensors*, 19(20):4503, 2019.